

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 15, 2006, 12:59:30 ; Search time 171 Seconds  
(without alignments)  
800.423 Million cell updates/sec

Title: US-10-751-235-1

Perfect score: 393

Sequence: 1 LQPVADGSAVMEAKFSQM.....STDLLPYWKIDALCKIVPRQ 77

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10751235/runat\_12012006\_120916\_2538/app\_query.fasta\_1.263  
-DB=Issued Patents NA -QWTF=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=ablosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10751235 @CGN 1 193 runat\_12012006\_120916\_2538 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/1 COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq.\*  
8: /cgn2\_6/ptodata/1/ina/RE COMB.seq.\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 200   | 50.9        | 1831   | 3  | US-08-948-564-15   |
| 2          | 196   | 49.9        | 404    | 3  | US-09-615-192A-156 |
| 3          | 196   | 49.9        | 404    | 3  | US-09-169-789-156  |
| 4          | 76    | 19.3        | 3150   | 3  | US-10-018-730A-3   |
| 5          | 74    | 18.8        | 1633   | 3  | US-09-583-447A-7   |
| 6          | 71.5  | 18.2        | 1349   | 3  | US-09-583-447A-5   |
| 7          | 71.5  | 18.2        | 1515   | 3  | US-09-583-447A-3   |
| 8          | 71.5  | 18.2        | 1659   | 3  | US-09-583-447A-1   |
| 9          | 71.5  | 18.2        | 1973   | 3  | US-09-583-447A-11  |

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

Sequence 156, App

Sequence 156, App

Sequence 3, Appl

Sequence 7, Appl

Sequence 5, Appl

Sequence 3, Appl

Sequence 11, Appl

Sequence 15, Appl

```

; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1831 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1747
US-08-948-564-15

Alignment Scores:
Pred. No.: 2,576-20 Length: 1831
Score: 200.00 Matches: 40
Percent Similarity: 73.24% Conservativeness: 12
Best Local Similarity: 56.34% Mismatches: 19
Query Match: 50.89% Indels: 0
DB: 3 Gaps: 0

US-10-751-235-1 (1-77) x US-08-948-564-15 (1-1831)
Qy 7 AspGlySerAlaValAsnMetGluAlaLysPheSerGlnMetThrLeuAspValIleGlyLeuSerLeuPhe 26
Db 647 GACTCAATTGAATTCGATGACCTTCTGTCGACCAAGATCCAGTTATTAAAGCA 706

Qy 27 LeuSerLeuPheAsnTyrAsnPheAspSerLeuThrThrAspSerProValIleGluAla 46
Db 707 CTTCGTGTTCACTATGACCTTCTGTCGACCAAGATCCAGTTATTAAAGCA 766

Qy 47 ValTyrThrAlaLeuLysGluAlaGluLeuArgSerThrAspLeuLeuProTyrTrpLys 66
Db 767 GTCTATGGCACTCTTTTGAAGCTGAACACAGATCCACTTCTACATTCATATTGGAAA 826

Qy 67 IleAspAlaLeuCysLysIleValProArgGln 77
Db 827 ATTCCATTGGCAAGGTGGATAGTCCCAAGGCAA 859

RESULT 2
US-09-615-192A-156
; Sequence 156, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003C4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 156
; LENGTH: 404
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-615-192A-156

Alignment Scores:
Pred. No.: 1,266-20 Length: 404
Score: 196.00 Matches: 39
Percent Similarity: 74.63% Conservativeness: 11
Best Local Similarity: 58.21% Mismatches: 17
Query Match: 49.87% Indels: 0
DB: 3 Gaps: 0

US-10-751-235-1 (1-77) x US-09-615-192A-156 (1-404)
Qy 11 ValAsnMetGluAlaLysPheSerGlnMetThrLeuAspValIleGlyLeuSerLeuPhe 30
Db 31 GTTGACATGGAGGAGCAATTTTCAAACTCTAGCTTTGGACATTATTGGATTGTGTATTT 90

Qy 31 AsnTyrAsnPheAspSerLeuThrThrAspSerProValIleGluAlaValTyrThrAla 50
Db 91 AACTATGATTTTGGATTCGGTACTCGAGAAATCACCAGTAATCAAGCAGTCTATGGTACA 150

Qy 51 LeuLysGluAlaGluLeuArgSerThrAspLeuLeuProTyrTrpLysIleAspAlaLeu 70
Db 151 TTGTTTGAAGCTGAGCATAGATCAACCTTTTACATACCATCTGGAATTTCCGCTGGCA 210

Qy 71 CysLysIleValProArgGln 77
Db 211 AGATGTTAGTTCTCTCGCCAA 231

RESULT 3
US-09-169-789-156
; Sequence 156, Application US/09169789
; Patent No. 6653528
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003C2
; CURRENT APPLICATION NUMBER: US/09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 156
; LENGTH: 404
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-169-789-156

Alignment Scores:
Pred. No.: 1,266-20 Length: 404
Score: 196.00 Matches: 39
Percent Similarity: 74.63% Conservativeness: 11
Best Local Similarity: 58.21% Mismatches: 17
Query Match: 49.87% Indels: 0
DB: 3 Gaps: 0

US-10-751-235-1 (1-77) x US-09-169-789-156 (1-404)
Qy 11 ValAsnMetGluAlaLysPheSerGlnMetThrLeuAspValIleGlyLeuSerLeuPhe 30
Db 31 GTTGACATGGAGGAGCAATTTTCAAACTCTAGCTTTGGACATTATTGGATTGTGTATTT 90

Qy 31 AsnTyrAsnPheAspSerLeuThrThrAspSerProValIleGluAlaValTyrThrAla 50
Db 91 AACTATGATTTTGGATTCGGTACTCGAGAAATCACCAGTAATCAAGCAGTCTATGGTACA 150

Qy 51 LeuLysGluAlaGluLeuArgSerThrAspLeuLeuProTyrTrpLysIleAspAlaLeu 70
Db 151 TTGTTTGAAGCTGAGCATAGATCAACCTTTTACATACCATCTGGAATTTCCGCTGGCA 210

Qy 71 CysLysIleValProArgGln 77
Db 211 AGATGTTAGTTCTCTCGCCAA 231

RESULT 4
US-10-018-730A-3
; Sequence 3, Application US/10018730A
; Patent No. 6794168
; GENERAL INFORMATION:

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 12, 2006, 20:19:34 ; Search time 12340 Seconds  
(without alignments)  
11364.084 Million cell updates/sec

Title: US-10-751-235-5  
Perfect score: 2467  
Sequence: 1 agggagcttcactcttttc.....tgaagtgagcaaggttaa 2467

Scoring table: IDENTITY NUC  
Gap 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl.\*  
1: gb.ba.\*  
2: gb.in.\*  
3: gb.env.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pr.\*  
9: gb.ro.\*  
10: gb.ste.\*  
11: gb.sy.\*  
12: gb.un.\*  
13: gb.vi.\*  
14: gb.htg.\*  
15: gb.pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description  |
|------------|-------|-------------|--------|----|--------------|
| 1          | 2467  | 100.0       | 92611  | 15 | AT4D2        |
| 2          | 803   | 32.5        | 1620   | 15 | AY424805     |
| 3          | 803   | 32.5        | 1858   | 15 | AY091083     |
| 4          | 803   | 32.5        | 3304   | 15 | AF367289     |
| 5          | 376.4 | 15.3        | 378    | 15 | AT526463     |
| 6          | 268.6 | 10.9        | 1077   | 15 | AK220829     |
| 7          | 252   | 10.2        | 1967   | 15 | BT012891     |
| 8          | 239.6 | 9.7         | 110000 | 15 | AP008216_208 |
| 9          | 239.6 | 9.7         | 178022 | 15 | AC025783     |
| 10         | 239.6 | 9.7         | 299735 | 15 | AB017117     |
| 11         | 193   | 7.4         | 1876   | 15 | AK065689     |
| 12         | 146.8 | 6.0         | 205408 | 14 | AC155564     |
| 13         | 91    | 3.7         | 1788   | 15 | AY142017     |
| 14         | 91    | 3.7         | 2017   | 15 | AY058173     |
| 15         | 91    | 3.7         | 2057   | 15 | AY056446     |
| 16         | 83.4  | 3.4         | 4217   | 15 | AK068163     |
| 17         | 75.2  | 3.0         | 178657 | 14 | AC155503     |
| 18         | 74.8  | 3.0         | 110000 | 14 | CR954209_3   |

|    |      |     |        |    |              |                     |
|----|------|-----|--------|----|--------------|---------------------|
| 19 | 73.2 | 3.0 | 1831   | 15 | AF022457     | AF022457 Glycine m  |
| 20 | 72.8 | 3.0 | 1926   | 15 | PSCYTP450    | 249263 P. sativum m |
| 21 | 72   | 2.9 | 1886   | 15 | BT002582     | BT002582 Arabidops  |
| 22 | 72   | 2.9 | 1921   | 15 | AY062675     | AY062675 Arabidops  |
| 23 | 71.8 | 2.9 | 110000 | 14 | CR954213_4   | Continuation (5 of  |
| 24 | 71.4 | 2.9 | 110000 | 14 | CR954201_08  | Continuation (9 of  |
| 25 | 71.2 | 2.9 | 625    | 15 | AK221212     | AK221212 Arabidops  |
| 26 | 70.4 | 2.9 | 119420 | 15 | AC079041     | AC079041 Arabidops  |
| 27 | 67   | 2.7 | 2093   | 15 | AK100596     | AK100596 Oryza sat  |
| 28 | 67   | 2.7 | 2401   | 15 | AK066680     | AK066680 Oryza sat  |
| 29 | 66.4 | 2.7 | 181167 | 14 | AC155486     | AC155486 Zea mays   |
| 30 | 66.2 | 2.7 | 2038   | 15 | AY601887     | AY601887 Ginkgo bi  |
| 31 | 65.2 | 2.6 | 103973 | 15 | AP004028     | AP004028 Oryza sat  |
| 32 | 65.2 | 2.6 | 110000 | 15 | AP008208_351 | Continuation (352   |
| 33 | 65.2 | 2.6 | 145014 | 15 | AP004048     | AP004048 Oryza sat  |
| 34 | 59.4 | 2.4 | 197405 | 15 | ATCHRIV40    | AL161540 Arabidops  |
| 35 | 59.4 | 2.4 | 202860 | 15 | ATFCA2       | Z97337 Arabidops    |
| 36 | 59   | 2.4 | 120761 | 15 | AC148994     | AC148994 Medicago   |
| 37 | 59   | 2.4 | 109093 | 15 | AC124218     | AC124218 Medicago   |
| 38 | 58.2 | 2.4 | 110000 | 14 | CR954215_1   | Continuation (2 of  |
| 39 | 57.8 | 2.3 | 404    | 6  | BD224432     | BD224432 Materials  |
| 40 | 57.8 | 2.3 | 404    | 6  | AR216482     | AR216482 Sequence   |
| 41 | 57.8 | 2.3 | 404    | 6  | AR432884     | AR432884 Sequence   |
| 42 | 56.6 | 2.3 | 2269   | 15 | AF459441     | AF459441 Skeletone  |
| 43 | 55.2 | 2.2 | 207718 | 14 | AC155487     | AC155487 Zea mays   |
| 44 | 53.8 | 2.2 | 86539  | 15 | AP004485     | AP004485 Lotus cor  |
| 45 | 53   | 2.1 | 110000 | 1  | CP000001_30  | Continuation (31 o  |

ALIGNMENTS

|            |                             |  |          |     |        |                 |
|------------|-----------------------------|--|----------|-----|--------|-----------------|
| RESULT 1   | ATT4D2                      | Arabidopsis thaliana DNA chromosome 3, BAC clone T4D2.   | 92611 bp | DNA | linear | PLN 16-APR-2005 |
| LOCUS      | ATT4D2                      | Arabidopsis thaliana DNA chromosome 3, BAC clone T4D2.   | 92611 bp | DNA | linear | PLN 16-APR-2005 |
| DEFINITION | AL132958                    | Arabidopsis thaliana (thale cress)   |          |     |        |                 |
| ACCESSION  | AL132958.1                  | GI:6434245   |          |     |        |                 |
| VERSION    | AL132958.1                  | GI:6434245   |          |     |        |                 |
| KEYWORDS   | Arabidopsis thaliana        |  |          |     |        |                 |
| SOURCE     | Arabidopsis thaliana        |  |          |     |        |                 |
| ORGANISM   | Arabidopsis thaliana        |  |          |     |        |                 |
| REFERENCE  | 1                           | Nyakatura,G., Partmann,B., Dauner,D., Sterr,W., Holland,R., Weichselgartner,M., Mewes,H.W., Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.  |          |     |        |                 |
| AUTHORS    | 2                           | Unpublished  |          |     |        |                 |
| JOURNAL    | 2                           | (bases 1 to 92611)   |          |     |        |                 |
| REFERENCE  | EU                          | Arabidopsis sequencing, project.   |          |     |        |                 |
| AUTHORS    | Direct Submission           |  |          |     |        |                 |
| TITLE      | Submitted (21-DEC-1999)     | MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; |          |     |        |                 |
| JOURNAL    | http://www.genoscope.cns.fr | Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.   |          |     |        |                 |
| COMMENT    | Location/Qualifiers         | 1..92611   |          |     |        |                 |
| FEATURES   | source                      | /organism="Arabidopsis thaliana"   |          |     |        |                 |
|            |                             | /mol_type="genomic DNA"  |          |     |        |                 |
|            |                             | /db_xref="taxon:3702"  |          |     |        |                 |
|            |                             | /chromosome="3"  |          |     |        |                 |
|            |                             | /ecotype="Columbia"  |          |     |        |                 |
|            |                             | 2081..2646   |          |     |        |                 |
|            |                             | /gene="T4D2.10"  |          |     |        |                 |
|            |                             | join(2081..2434,2533..2646)  |          |     |        |                 |
|            |                             | /gene="T4D2.10"  |          |     |        |                 |
| gene       |                             |  |          |     |        |                 |
| CDS        |                             |  |          |     |        |                 |

```
/note="similarity to BETA-GALACTOSIDASE PRECURSOR,
Lycopodium sculentum, gb:P48980"
/codon_start=1
/product="putative protein"
/db_xref="GI:6630728"
/db_xref="GOA:Q9SCQ3"
/db_xref="InterPro:IPR000922"
/db_xref="UniProt/TREMBL:Q9SCQ3"
/translation="METLQCRHQHVFTLLVLHSSFLGLASKIDVSDARGIRIDGG
OKRFLTNSPQKGEAACTNEEPDLGPLTRISCNPGYVITKINFADYGNPTGTCGHF
RRGCGARATWRIVKNCCLGKCHLLVTDMEFGPSCKGAPMLAVETTCTIA"
2081..2434
/gene="T4D2.10"
/number=1
/number=1
/gene="T4D2.10"
/number=1
/number=1
2533..2646
/gene="T4D2.10"
/number=2
3341..8521
/gene="T4D2.20"
/number=1
join(3341..3364,3812..3934,4065..4305,4394..4956,
5131..6090,6168..6490,6606..6884,6997..7249,7359..7529,
7645..7737,7816..7923,8048..8164,8257..8346,8438..8521)
/gene="T4D2.20"
/note="similarity to KIAA0010 protein, Homo sapiens,
gb:A38919"
/codon_start=1
/product="putative protein"
/protein_id="CAB64211.1"
/db_xref="GI:6630729"
/db_xref="GOA:Q9SCQ2"
/db_xref="InterPro:IPR000048"
/db_xref="InterPro:IPR000569"
/db_xref="UniProt/TREMBL:Q9SCQ2"
/translation="MDLNRRKHVSLRGASSGEISRDALLAKVQERELRSVARRANAA
SLFTQVRWSYIVRKAATIEIOBENWLLSCHVTLTKSWSSRVLRFPLFFVRSLSV
QFKIQAREIHCMTQCFKILLESINSDQYNGFCSLAVGTSEDSKTWACOTRRVSLC
SHLTGECYSOERIKNVIGVNAALLRLILVLPDKSWKIITNENFEDAETAKTIIQF
IGSCKGYSAVRVYIKTLTKHTDERLVITTSVTLALRPHVKQAPVDDNQDPTNL
ABEYVSILITPLRVCVLPALRALKHSILMPSFTILLKDKILNITISENSE
KQSCMTGPIVGVNIGNIIISLATVSETDFDPQSNPEMFVLYVHVITLAEALLSQ
VESVGIQDHLIDIEATNETEGNSVKISFVEMLRPVCCQWHLAKLAAAGKEIRVA
DKDASTSGKSETGLIDIDARLYSCMLRIFCVNVPVLPVPLVLMLSFCPGYIVSLW
NSLESVLLPENGCTADDAHSGSANTSNTRSPSEKCLKHLKNDVKNWVNLKFSKG
SPGREHVECTSDQPGSQVNESNDVWDETLLRGPGVIGSKVSCILLHLFCATYHL
LVLLDDIQFYBKQVFFMLEKQORIASMLNTLVYGLLKGTPGSRQLMDSAIRCLHLL
YERDCRHPFCASALWLSFGRTSRPIAPARTHEVLTPTSDVLTFTPSGVSITITPHVP
PFERHVFREFFISKASRWAGEVAPGARSTEVVRGHVVEDGFQQLNSIGSRL
KSSITHVSFVNESGLPEAGLDYGLSKFELTDITKAAPATEYGLFSQTPSDRLAVPSP
SARHLENGIOMEPLGRIVGKALYEGILLDYSESHVPIOKLGRYSRIDELSGLDPEL
YRNLVYKHVGDGLKELCLDFTVTEFCGKMSIIELPFGKDTSVTNENKQYIHAMA
DYKLRQIVPSPNAPRGLTDLISPAWLKNAFHEFNQLLGGNHDIDVDDLRCTKY
TGGYSDSRTKIFWEMVKGFPEBSCILLKLFVTSCSRPLLGFYKLYQPTPIIHKVC
DTSLWAAIGGQDRELPSASTCYNTLKLPTTKRASITWREKLLYAITSNAGFELS"
3341..3364
/gene="T4D2.20"
/number=1
3365..3811
/gene="T4D2.20"
/number=1
3812..3934
/gene="T4D2.20"
/number=2
3935..4064
/gene="T4D2.20"
/number=2
4065..4305
/gene="T4D2.20"
/number=3
4306..4393
/gene="T4D2.30"
/number=1
/gene="T4D2.20"
/number=3
4394..4956
/gene="T4D2.20"
/number=4
4957..5130
/gene="T4D2.20"
/number=4
5131..6090
/gene="T4D2.20"
/number=5
6091..6167
/gene="T4D2.20"
/number=5
6168..6490
/gene="T4D2.20"
/number=6
6491..6605
/gene="T4D2.20"
/number=6
6606..6884
/gene="T4D2.20"
/number=7
6885..6996
/gene="T4D2.20"
/number=7
6997..7249
/gene="T4D2.20"
/number=8
7250..7358
/gene="T4D2.20"
/number=8
7359..7529
/gene="T4D2.20"
/number=9
7530..7644
/gene="T4D2.20"
/number=9
7645..7737
/gene="T4D2.20"
/number=10
7738..7815
/gene="T4D2.20"
/number=10
7816..7923
/gene="T4D2.20"
/number=11
7924..8047
/gene="T4D2.20"
/number=11
8048..8164
/gene="T4D2.20"
/number=12
8165..8256
/gene="T4D2.20"
/number=12
8257..8346
/gene="T4D2.20"
/number=13
8347..8437
/gene="T4D2.20"
/number=13
8438..8521
/gene="T4D2.20"
/number=14
9057..10601
/gene="T4D2.30"
complement(join(9057..9151,9403..9658,9810..10043,
10124..10254,10370..10601))
/gene="T4D2.30"
complement(join(9057..9151,9403..9658,9810..10043,
10124..10254,10370..10601))
/gene="T4D2.30"
CDS
```

```

/note="strong similarity to proline-rich protein, Raphanus
sativa, PIR:S16748"
/codon start=1
/product="putative protein"
/protein id="CAB64213_1"
/db xref="GI:6630730"
/db xref="GOA:Q9SCQ1"
/db xref="InterPro:IPR001087"
/db xref="UniProt:TREMBL:Q9SCQ1"
/translation="NRVSQFVLLVLSQFCKSGKGVAVPALIMFGDSIVDVGNNNLL
SI VKSNFLPVGRDIDIDQPTGFCFCKGLAVDFSAEYLGFSYPPAFVREASNNELLI
CANFASASGGYDFTASVFGSISLTLRQSYTYRAQNRVTMYGRNARILFSGRTHIL
SAGSDFLQNYINPLMLITDQDFALLRSEPTIQLVGGARRGVISLSPMG
CPAAATLFGAGNKSCHPTEAWEALLCOLLGGISILIN"
IITPTDNGVFWDFGCVPTAWEALLCOLLGGISILIN"

```

[illegible]

|   |   |
|---|---|
| RESULT 2  |   |
| AY424805  |   |
| LOCUS   |   |
| DEFINITION  | linear PLN 07-JAN-2004  |
|   | Arabidopsis thaliana chloroplast carotenoid epsilon-ring            |
|   | hydroxylase (LUT1) mRNA, complete cds; nuclear gene for chloroplast |
| ACCESSION   |   |
| AY424805  |   |
| KEYWORDS  |   |
| SOURCE  |   |
| ORGANISM  |   |
|   | Arabidopsis thaliana (thale cress)                                  |
|   | Arabidopsis thaliana  |
|   | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  |
|   | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  |
|   | rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.        |
| REFERENCE   |   |
| AUTHORS   |   |
|   | 1 (bases 1 to 1620)   |
|   | Tian,L., Musetti,V., Kim,J., Magallanes-Lundback,M. and             |
|   | DellaPenna,D.   |
| TITLE   |   |
|   | The Arabidopsis LUT1 locus encodes a member of the cytochrome P450  |
|   | family that is required for carotenoid epsilon-ring hydroxylation   |
|   | activity  |
| JOURNAL   |   |
| PUBLISHED   |   |
| PROC. NATL. ACAD. SCI. U.S.A. 101 (1), 402-407 (2004)       |   |
| 14709673  |   |
| 2 (bases 1 to 1620)   |   |
| Tian,L., Musetti,V., Kim,J., Magallanes-Lundback,M. and     |   |
| DellaPenna,D.   |   |
| Direct Submission   |   |
| Submitted (30-SEP-2003) Biochemistry and Molecular Biology, |   |
| Michigan State University, East Lansing, MI 48824, USA      |   |
| Location/Qualifiers   |   |
| 1. 1620   |   |
| FEATURES  |   |
| source  |   |

[illegible]